

Penerapan Data Mining Untuk Memprediksi Prestasi Akademik Mahasiswa Menggunakan Algoritma C4.5 dengan CRISP-DM

Suprayuandi Pratama ^{1*}, Iswandi ^{2*}, Andre Sevtian ^{3*} Tsabita Putri Anjani ^{4*}

* Ilmu Komputer, Universitas Muhammadiyah Bangka Belitung

suprayuandi.pratama@unmuhbabel.ac.id ¹, iswandi@unmuhbabel.ac.id ², andresevtian501@gmail.com ³, tsabitha266@gmail.com ⁴

Article Info

Article history:

Received 2022-12-22

Revised 2023-01-19

Accepted 2023-01-24

Keyword:

Data Mining,
Prestasi Akademik,
Algoritma C4.5,
CRISP-DM,

ABSTRACT

Student data is crucial, particularly when it comes to academic performance as measured by the student achievement index. The information now available is only a data mine that has been organized and saved in the system. This occurred at Muhammadiyah University of Bangka Belitung's Faculty of Engineering and Science, where student data was merely a database and not effectively used. This information needs to be examined in order to be categorized based on factors including the student achievement index, residence, discipline, and the lecturer's position. To collect the necessary dataset for this project, questionnaires will be distributed, and CRISP-DM will be used in data mining techniques. And the C4.5 method, which makes use of the RapidMiner application, is the algorithm used in the classification process. 85 datasets are categorised, and they produce good results with a 67.06% accuracy rate.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Teknologi basis data sangat berkembang pesat maka kebutuhan informasi yang lebih dari sebuah data yang tersimpan dalam database yaitu pengetahuan (*knowledge*). Informasi yang dibutuhkan tidak bisa diperoleh dengan mudah sebab volume data yang besar dan kompleks [1]. Dibutuhkan sebuah metode untuk mendapatkan pengetahuan yang tidak bisa terlihat di dalam data namun sangat potensial untuk digunakan yaitu metode data mining. Pada data mahasiswa diterapkan sebuah metode data mining klasifikasi dengan Algoritma C4.5 untuk memprediksi prestasi akademik mahasiswa. Data mining adalah proses menemukan pola yang menarik dan pengetahuan dari sejumlah besar data. Sumber data dapat mencakup database, data gudang, web, repositori informasi lain, atau data yang dialirkan kesistem secara dinamis [1] [2]. Data mining merupakan proses menemukan pengetahuan yang menarik dari data yang berjumlah besar yang disimpan di dalam database, gudang data atau repositori informasi. Data mining juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data. Algoritma C4.5 merupakan kelompok algoritma decision tree [3]. Algoritma ini mempunyai input berupa *training samples* dan *samples*. *Training samples* berupa data

contoh yang akan digunakan untuk membangun sebuah *tree* yang telah diuji kebenarannya. Sedangkan *samples* merupakan *field-field* data yang nantinya akan kita gunakan sebagai parameter dalam melakukan klasifikasi data. Algoritma C4.5 adalah algoritma hasil pengembangan dari algoritma ID3.

Fakultas Teknik dan Sains Universitas Muhammadiyah Bangka Belitung [4] terus berupaya agar mencapai tujuan pendidikan yang berkemajuan dan mencerahkan serta berkualitas agar kemampuan professional mahasiswa di bidangnya masing-masing. Salah satu indikator untuk mencapai tujuan tersebut adalah hasil prestasi akademik mahasiswa yang dinyatakan dengan Indek Prestasi Kumulatif (IPK) dan Indek Prestasi Semester (IPS). Indek Prestasi Semester merupakan hasil nilai akademik mahasiswa pada mata kuliah yang telah ditempuh untuk di tiap semesternya. dan Indek Prestasi Kumulatif merupakan prestasi akademik mahasiswa yang menggabungkan semua hasil nilai mata kuliah yang telah ditempuh.

Untuk mewujudkan pendidikan yang berkualitas, berkemajuan dan mencerahkan berkaitan dengan peran dosen, kedisiplinan mahasiswa, sosial ekonomi mahasiswa dan juga peran dosen terhadap hasil belajar mahasiswa. Hal tersebut di buktikan dengan banyaknya penelitian yang telah dilakukan sebelumnya. Fakultas Teknik dan Sains (FTS) Universitas

Muhammadiyah Bangka Belitung (Unmuh Babel) mempunyai database akademik mahasiswa. Basisdata ini berpotensi untuk menghasilkan sebuah informasi baru yang sangat berguna. Salah satu yang bisa dilakukan oleh data mining adalah memprediksi prestasi akademik [5] para mahasiswa Fakultas Teknik dan Sains. Jika prestasi akademik mahasiswa dapat di ketahui lebih dini maka pihak program studi dapat melakukan tindakan yang diperlukan agar mahasiswa dapat mencapai prestasi akademik yang baik. Data mining prestasi mahasiswa tersebut perlunya diklasifikasikan untuk menghasilkan sebuah keputusan. Pengklasifikasian data mining ini menggunakan teknik algoritma C4.5 (*Decision Tree*) dan juga menggunakan metode CRISP-DM (Cross Industry Standard Process for Data Mining) [6]. Aplikasi yang digunakan untuk pengklasifikasian ini yaitu Rapid Miner dengan menerapkan untuk Algoritma C4.5

Beberapa penelitian sebelumnya telah melakukan mining terhadap data pendidikan [7] sehingga memudahkan dalam menganalisa aktifitas dari institusi pendidikan. Kemudian algoritma C4.5 diterapkan dalam mengukur tingkat kepuasan mahasiswa [8] serta prediksi kelulusan mahasiswa [9] menggunakan metode CRISP-DM [10] [6]. Pada penelitian kali ini adalah menerapkan data mining untuk memprediksi prestasi akademik mahasiswa dengan menggunakan teknik Algoritma C4.5 (*decision tree*) dengan mengklasifikasikan berdasarkan peran dosen, kedisiplinan mahasiswa, sosial ekonomi mahasiswa dan juga hasil belajar masa lalu. Tujuan dari penelitian ini yaitu untuk mengklasifikasikan prestasi akademi mahasiswa dan menghasilkan sebuah keputusan sehingga Fakultas Teknik dan Sains bisa melakukan langkah yang baik untuk meningkatkan prestasi akademik mahasiswa.

II. METODE

Penelitian ini dilaksanakan di Fakultas Teknik dan Sains Universitas Muhammadiyah Bangka Belitung. Waktu penelitian dilaksanakan pada mulai bulan Juli 2022. Target penelitian ini adalah mahasiswa Fakultas Teknik dan Sains Universitas Muhammadiyah Bangka Belitung. Pada penelitian ini menggunakan 2 metode/pendekatan yaitu Metode Pengumpulan Data dan Metode CRISP-DM (Cross Industry Standard Process for Data Mining) untuk proses dalam Data Mining nya.

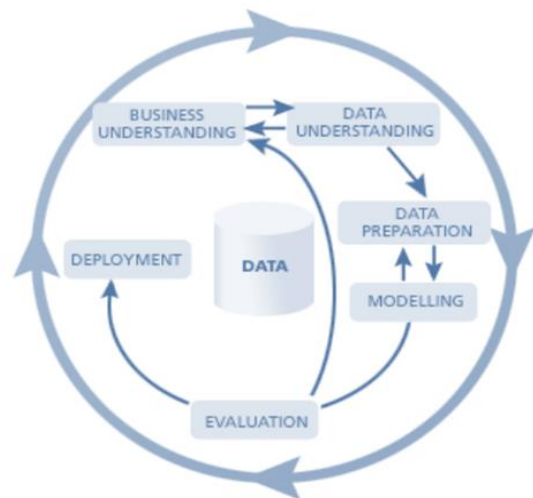
A. Teknik Pengumpulan data

Teknik pengumpulan data yang digunakan untuk penelitian ini adalah observasi dan kuesioner. Observasi terhadap data mahasiswa yang ada pada sistem akademik untuk mendapatkan indeks prestasi mahasiswa pada program studi yang ada di Fakultas Teknik dan Sains. Penyebaran kuisoner kepada mahasiswa di Fakultas Teknik dan Sains untuk mendapatkan data terkait tempat tinggal, kedisiplinan mahasiswa, dan peran dosen dalam perkuliahan. Dalam penyebaran kuisoner ini menggunakan Google Formulir.

B. Metode CRISP-DM (Cross Industry Standard Process for Data Mining)

Penelitian ini menggunakan metode CRISP-DM dalam proses Data Mining. Metode ini terdiri dari 6 tahapan:

- Pemahaman terhadap bisnis (Business/Research Understanding Phase) dimana data yang diperoleh dari database akademik Mahasiswa Fakultas Teknik dan Sains selama ini belum pernah dilakukan penggalian kekayaan terhadap data tersebut
- Fase Pemahaman Data (Data Understanding Phase) Tahapan ini diawali dengan mengumpulkan data menggunakan kuisoner, mendeskripsikan data, serta mengevaluasi kualitas data.
- Persiapan data (Data Preparation) Dari hasil data mahasiswa maka akan dilakukan teknik Data Preparation agar kualitas data diperoleh lebih baik.
- Fase Pemodelan (Modeling Phase) Pada tahapan ini merupakan tahapan pemrosesan dataset yang diklasifikasikan oleh model dan menentukan akan dari dataset
- Fase Evaluasi (Evaluation Phase) Pada fase ini dilakukan pengujian terhadap model klasifikasi yang telah di modelkan pada fase pemodelan.
- Fase Penyebaran (Deployment Phase) Setelah pembentukan model selanjutnya dapat digunakan untuk memprediksi prestasi akademik mahasiswa oleh Program Studi di lingkungan Fakultas Teknik dan Sains.



Gambar 1. Metode CRISP-DM (Cross Industry Standard Process for Data Mining)

C. Metode Data Mining Menggunakan Algoritma C4.5

Secara garis besar langkah-langkah yang dilakukan oleh algoritma C4.5 dalam membentuk pohon keputusan adalah sebagai berikut:

- 1) Pada saat awal pembentukan pohon akan dimulai dengan membuat suatu node yang melambangkan *training sample*.

- 2) Jika sampel-sampel tersebut memiliki kelas yang sama, maka node tersebut dijadikan *leaf node* dengan label kelas tersebut.
- 3) Jika sampel-sampel tersebut tidak memiliki kelas yang sama maka algoritma ini akan mencari *gain ratio* tertinggi dari atribut yang tersedia, sebagai cara untuk memilih atribut yang paling berpengaruh pada training sample yang tersedia. Nantinya atribut ini akan dijadikan atribut “penguji” atau “keputusan” pada node tersebut. Hal yang perlu diperhatikan adalah ketika atribut tersebut bernilai kontinu, maka atribut tersebut harus di diskritkan terlebih dahulu.
- 4) Cabang untuk setiap *node* akan dibentuk berdasarkan nilai-nilai yang diketahui dari atribut pengujian.
- 5) Algoritma ini akan terus melakukan proses yang sama rekursif untuk membentuk suatu pohon keputusan untuk setiap sample di setiap bagiannya.
- 6) Proses rekursif ini akan berhenti, ketika salah satu dari kondisi di bawah telah terpenuhi. Kondisi-kondisinya adalah:
 - Semua sample yang diberikan pada *node* adalah berasal dari satu kelas yang sama.
 - Tidak ada atribut lainnya yang dapat digunakan untuk mempartisi sample lebih lanjut.
 - Tidak ada sample yang memenuhi $test-attribute = a_i$
 Dalam hal ini, sebuah daun dibuat dan dilabeli dengan kelas yang memiliki sample terbanyak (*majority voting*).

III. HASIL DAN PEMBAHASAN

Data yang peneliti gunakan dalam penelitian ini adalah data kuesioner. Data diperoleh dari penyebaran kuesioner sebanyak 85 responden Fakultas Teknik dan Sains angkatan 2020/2021 dan angkatan 2021/2022.

A. Pemahaman terhadap bisnis (*Business/Research Understanding Phase*)

Penerapan data mining pada penelitian ini berhubungan langsung dengan data mahasiswa Fakultas Teknik dan Sains untuk menggali prestasi mahasiswa, kedisiplinan dan peran dosen dalam perkuliahan.

B. Fase Pemahaman Data (*Data Understanding Phase*)

Data yang digunakan untuk melakukan perhitungan menggunakan metode Algoritma C4.5 di dapatkan pada database akademi dan juga kuisoner yang telah di isi oleh mahasiswa. Dari kuisoner yang disebar pada mahasiswa Fakultas Teknik dan Sains didapatkan 85 ExampleSet yang dapat dilihat pada tabel di bawah ini.

TABEL.1
DATASET HASIL KUISONER

No	Fakultas Teknik dan Sains Program Studi	Indeks Prestasi Semester 1	Indeks Prestasi Semester 2	Tempat Tinggal	Peran Dosen dalam Perkuliahan	Kedisiplinan Mahasiswa dalam Perkuliahan
1	ILKOM	3.41	3.48	Rumah Sendiri	2	3
2	ILKOM	3.64	3.71	Rumah Sendiri	3	3
3	ILKOM	3.64	3.6	Kos	3	3
4	ILKOM	3.61	3.75	Kos	3	3
5	ILKOM	3.27	3.36	Kos	3	3
6	ILKOM	3.43	3.3	Kos	4	4
7	ILKOM	3.44	3.26	Kos	2	1
8	ILKOM	3.52	3.32	Kos	4	4
9	KSDA	3.43	3.35	Kos	4	3
10	KSDA	3.59	3.34	Rumah Sendiri	4	3
11	KSDA	3.29	3.22	Rumah Sendiri	4	4
12	KSDA	3.63	3.65	Kos	3	1
13	KSDA	3.9	3.75	Rumah Sendiri	3	2
14	Sipil	3.48	3.57	Rumah Sendiri	4	3
15	Sipil	3.40	3.47	Rumah Sendiri	4	4
16	Sipil	3.48	3.57	Rumah Sendiri	3	3
17	Sipil	3.40	3.47	Kos	4	4
18	Sipil	3.72	3.71	Rumah Sendiri	4	3
19	Sipil	3.30	3.48	Kos	4	4
20	Sipil	3.43	3.5	Rumah Sendiri	4	4
...
85	Sipil	3.33	3.32	Kos	4	4

C. Persiapan data (*Data Preparation*)

Pada Tahap Data Cleaning, dataset yang terdiri dari 85 data yang akan di analisis menggunakan Rapid Miner Decision Tree menggunakan Algoritma C4.5. Berikut data yang sudah di persiapkan.

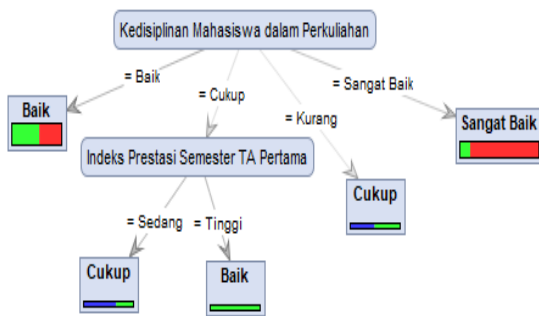
TABEL.2
DATA PERSIAPAN

No	Indeks Prestasi Kumulatif	Tempat Tinggal	Peran Dosen dalam Perkuliahan	Kedisiplinan Mahasiswa dalam Perkuliahan
1	Sedang	Rumah Sendiri	Cukup	Baik
2	Tinggi	Rumah Sendiri	Baik	Baik
3	Tinggi	Kos	Baik	Baik
4	Kurang	Kos	Sangat Baik	Sangat Baik
5	Tinggi	Kos	Sangat Baik	Sangat Baik
6	Tinggi	Kos	Sangat Baik	Sangat Baik

7	Tinggi	Rumah Sendiri	Sangat Baik	Sangat Baik
8	Sedang	Kos	Baik	Baik
9	Tinggi	Kos	Baik	Baik
10	Tinggi	Rumah Sendiri	Sangat Baik	Baik
11	Tinggi	Rumah Sendiri	Sangat Baik	Sangat Baik
12	Sedang	Kos	Sangat Baik	Baik
13	Tinggi	Kos	Baik	Baik
14	Sedang	Rumah Sendiri	Baik	Baik
15	Sedang	Kos	Baik	Sangat Baik
16	Tinggi	Rumah Sendiri	Sangat Baik	Sangat Baik
17	Kurang	Kos	Sangat Baik	Baik
18	Tinggi	Kos	Baik	Baik
19	Sedang	Kos	Baik	Baik
20	Sedang	Kos	Sangat Baik	Sangat Baik
...
	Sedang	Kos	Cukup	Kurang

D. Fase Pemodelan (Modeling Phase)

Pada tahapan ini merupakan tahapan pemrosesan dataset persiapan yang diklasifikasikan oleh model dan kemudian menghasilkan sejumlah aturan. Pada penelitian ini menggunakan algoritma C 4.5. Dari hasil proses data mining tersebut diperoleh rule atau aturan sbb:



Gambar 2. Hasil dari Decision Tree

Dari hasil pohon keputusan diatas didapatkan deskripsinya sebagai berikut

- Kedisiplinan Mahasiswa dalam Perkuliahan = Baik: Baik {Cukup=1, Baik=26, Sangat Baik=21}
- Kedisiplinan Mahasiswa dalam Perkuliahan = Cukup
- Indeks Prestasi Semester TA Pertama = Sedang: Cukup {Cukup=2, Baik=1, Sangat Baik=0}
- Indeks Prestasi Semester TA Pertama = Tinggi: Baik {Cukup=0, Baik=3, Sangat Baik=0}
- Kedisiplinan Mahasiswa dalam Perkuliahan = Kurang: Cukup {Cukup=1, Baik=1, Sangat Baik=0}
- Kedisiplinan Mahasiswa dalam Perkuliahan = Sangat Baik: Sangat Baik {Cukup=0, Baik=4, Sangat Baik=25}.

E. Fase Evaluasi (Evaluation Phase)

Pada fase ini dilakukan pengujian terhadap model klasifikasi dengan algoritma C4.5. Pengujian ini dilakukan dengan analisis tingkat Performansinya. Berikut hasil pengujiannya pada data set kuisoner

accuracy: 67.06%

	true Cukup	true Baik	true Sangat Baik	class precision
pred. Cukup	3	2	0	60.00%
pred. Baik	1	29	21	56.86%
pred. Sangat Baik	0	4	25	86.21%
class recall	75.00%	82.86%	54.35%	

Gambar 3. Hasil Pengujian Data Set

Dari hasil pengujian di atas di dapatkan accuracy sebesar 67.06% dengan class Precision predikat cukup=60%, baik=56.86% dan sangat baik 86.21% dan class recall cukup =75%, baik=82.86%, sangat baik= 82.86%.

F. Deployment

Setelah tahap evaluasi dimana menilai secara detail hasil dari sebuah model maka dilakukan pengimplementasian dari keseluruhan model yang telah dibangun. Selain itu juga dilakukan penyesuaian terhadap model sehingga dapat menghasilkan suatu hasil yang sesuai dengan target awal tahap CRISP-DM ini.

IV. KESIMPULAN

Pada penelitian ini, telah dilakukan klasifikasi data mahasiswa Fakultas Teknik dan Sains menggunakan metode decision tree algoritma C4.5 dengan teknik data mining CRISP-DM. Dataset terdiri dari indeks prestasi mahasiswa, tempat tinggal, kedisiplinan, peran dosen dalam perkuliahan. Dari 85 Dataset telah mendapatkan sebuah keputusan terhadap prestasi akademik mahasiswa di Fakultas Teknik dan Sains Universitas Muhammadiyah Bangka Belitung meliputi Program Studi Ilmu Komputer, Program Studi Teknik Sipil dan Program Studi Konservasi Sumber Daya Alam serta mendapatkan hasil performa dari hasil accuracy sebesar 67.06% dengan class Precision predikat cukup=60%, baik=56.86% dan sangat baik 86.21% dan class recall cukup =75%, baik=82.86%, sangat baik= 82.86%. Hal ini menjadi acuan para dosen dan Mahasiswa untuk meningkatkan proses pembelajaran dalam perkuliahan serta kedisiplinan mahasiswa untuk berkuliah.

DAFTAR PUSTAKA

[1] J. Han, 'Data Mining: Concepts and Techniques Second Edition', Jan. 2023.
 [2] P.-N. Tan, M. Steinbach, and V. Kumar, 'Data mining cluster analysis: basic concepts and algorithms', *Introd. to data Min.*, pp. 487–533, 2013.
 [3] W. Crisnawaty Manalu, A. Muhazir, D. Setiawan, S. Informasi, S. Triguna Dharma, and T. Komputer, 'Penerapan Data Mining Untuk Memprediksi Minat

- Masyarakat Terhadap Asuransi Jiwa Dengan Metode Algoritma C4.5', *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 2, no. 1, pp. 169–178, Jan. 2023.
- [4] 'Visi, Misi & Tujuan - Universitas Muhammadiyah Bangka Belitung'. [Online]. Available: <https://unmuhbabel.ac.id/visi-misi-tujuan/>. [Accessed: 28-Feb-2023].
- [5] P. Prasetyawan, I. Ahmad, R. I. Borman, Ardiansyah, Y. A. Pahlevi, and D. E. Kurniawan, 'Classification of the Period Undergraduate Study Using Back-propagation Neural Network', in *2018 International Conference on Applied Engineering (ICAE)*, 2018, pp. 1–5.
- [6] M. A. Hasanah, S. Soim, and A. S. Handayani, 'Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir', *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 103–108, Oct. 2021.
- [7] S. A. A. Kharis and A. H. A. Zili, 'Learning Analytics dan Educational Data Mining pada Data Pendidikan', *J. Ris. Pembelajaran Mat. Sekol.*, vol. 6, no. 1, pp. 12–20, Mar. 2022.
- [8] N. Tulus Ujjianto and N. A. Ramdhan, 'Implementasi Data Mining C4.5 dalam Mengukur Tingkat Kepuasan Mahasiswa terhadap Pelayanan Akademik', *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 4, no. 01, pp. 115–111, May 2022.
- [9] T. H. Hasibuan and D. Mahdiana, 'Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 Pada Uin Syarif Hidayatullah Jakarta', *SKANIKA Sist. Komput. dan Tek. Inform.*, vol. 6, no. 1, pp. 61–74, Jan. 2023.
- [10] J. A. Solano, D. J. Lancheros Cuesta, S. F. Umaña Ibáñez, and J. R. Coronado-Hernández, 'Predictive models assessment based on CRISP-DM methodology for students performance in Colombia - Saber 11 Test', *Procedia Comput. Sci.*, vol. 198, pp. 512–517, Jan. 2022.